A simplified approach using Taqman low-density array for medulloblastoma subgrouping

Cruzeiro, Gustavo Alencastro Veiga ; Salomão, Karina Bezerra ; de Biagi, Carlos Alberto Oliveira ; Baumgartner, Martin ; Sturm, Dominik ; Lira, Régia Caroline Peixoto ; de Almeida Magalhães, Taciani ; Baroni Milan, Mirella ; da Silva Silveira, Vanessa ; Saggioro, Fabiano Pinto ; de Oliveira, Ricardo Santos ; Dos Santos Klinger, Paulo Henrique ; Seidinger, Ana Luiza ; Yunes, José Andrés ; de Paula Queiroz, Rosane Gomes ; Oba-Shinjo, Sueli Mieko ; Scrideli, Carlos Alberto ; Nagahashi, Suely Marie Kazue ; Tone, Luiz Gonzaga ; Valera, Elvis Terci

Abstract: Next-generation sequencing platforms are routinely used for molecular assignment due to their high impact for risk stratification and prognosis in medulloblastomas. Yet, low and middle-income countries still lack an accurate cost-effective platform to perform this allocation. TaqMan Low Density array (TLDA) assay was performed using a set of 20 genes in 92 medulloblastoma samples. The same methodology was assessed in silico using microarray data for 763 medulloblastoma samples from the GSE85217 study, which performed MB classification by a robust integrative method (Transcriptional, Methylation and cytogenetic profile). Furthermore, we validated in 11 MBs samples our proposed method by Methylation Array 450 K to assess methylation profile along with 390 MB samples (GSE109381) and copy number variations. TLDA with only 20 genes accurately assigned MB samples into WNT, SHH, Group 3 and Group 4 using Pearson distance with the average-linkage algorithm and showed concordance with molecular assignment provided by Methylation Array 450 k. Similarly, we tested this simplified set of gene signatures in 763 MB samples and we were able to recapitulate molecular assignment with an accuracy of 99.1% (SHH), 94.29% (WNT), 92.36% (Group 3) and 95.40% (Group 4), against 97.31, 97.14, 88.89 and 97.24% (respectively) with the Ward.D2 algorithm. t-SNE analysis revealed a high level of concordance (k = 4) with minor overlapping features between Group 3 and Group 4. Finally, we condensed the number of genes to 6 without significantly losing accuracy in classifying samples into SHH, WNT and non-SHH/non-WNT subgroups. Additionally, we found a relatively high frequency of WNT subgroup in our cohort, which requires further epidemiological studies. TLDA is a rapid, simple and cost-effective assay for classifying MB in low/middle income countries. A simplified method using six genes and restricting the final stratification into SHH, WNT and non-SHH/non-WNT appears to be a very interesting approach for rapid clinical decision-making.

DOI: https://doi.org/10.1186/s40478-019-0681-y

Originally published at:
Cruzeiro, Gustavo Alencastro Veiga; Salomão, Karina Bezerra; de Biagi, Carlos Alberto Oliveira; Baumgartner, Martin; Sturm, Dominik; Lira, Régia Caroline Peixoto; de Almeida Magalhães, Taciani; Baroni Milan, Mirella; da Silva Silveira, Vanessa; Saggioro, Fabiano Pinto; de Oliveira, Ricardo Santos; Dos Santos Klinger, Paulo Henrique; Seidinger, Ana Luiza; Yunes, José Andrés; de Paula Queiroz, Rosane Gomes; Oba-Shinjo, Sueli Mieko; Scrideli, Carlos Alberto; Nagahashi, Suely Marie Kazue; Tone, Luiz Gonzaga; Valera, Elvis Terci (2019). A simplified approach using Taqman low-density array for medulloblastoma subgrouping. Acta Neuropathologica Communications, 7:33.
DOI: https://doi.org/10.1186/s40478-019-0681-y

Cruzeiro et al. Acta Neuropathologica Communications (2019) 7:33
https://doi.org/10.1186/s40478-019-0681-y

Acta Neuropathologica Communications

RESEARCH Open Access

A simplified approach using Taqman low-density array for medulloblastoma subgrouping

Gustavo Alencastro Veiga Cruzeiro[1*], Karina Bezerra Salomão[1], Carlos Alberto Oliveira de Biagi Jr[2], Martin Baumgartner[3], Dominik Sturm[4], Régia Caroline Peixoto Lira[1], Taciani de Almeida Magalhães[2], Mirella Baroni Milan[2], Vanessa da Silva Silveira[2], Fabiano Pinto Saggioro[5], Ricardo Santos de Oliveira[6], Paulo Henrique dos Santos Klinger[1], Ana Luiza Seidinger[7], José Andrés Yunes[7], Rosane Gomes de Paula Queiroz[1], Sueli Mieko Oba-Shinjo[8], Carlos Alberto Scrideli[1], Suely Marie Kazue Nagahashi[8], Luiz Gonzaga Tone[1†] and Elvis Terci Valera[1†]

Abstract

Next-generation sequencing platforms are routinely used for molecular assignment due to their high impact for risk stratification and prognosis in medulloblastomas. Yet, low and middle-income countries still lack an accurate cost-effective platform to perform this allocation. TaqMan Low Density array (TLDA) assay was performed using a set of 20 genes in 92 medulloblastoma samples. The same methodology was assessed in silico using microarray data for 763 medulloblastoma samples from the GSE85217 study, which performed MB classification by a robust integrative method (Transcriptional, Methylation and cytogenetic profile). Furthermore, we validated in 11 MBs samples our proposed method by Methylation Array 450 K to assess methylation profile along with 390 MB samples (GSE109381) and copy number variations. TLDA with only 20 genes accurately assigned MB samples into WNT, SHH, Group 3 and Group 4 using Pearson distance with the average-linkage algorithm and showed concordance with molecular assignment provided by Methylation Array 450 k. Similarly, we tested this simplified set of gene signatures in 763 MB samples and we were able to recapitulate molecular assignment with an accuracy of 99.1% (SHH), 94.29% (WNT), 92.36% (Group 3) and 95.40% (Group 4), against 97.31, 97.14, 88.89 and 97.24% (respectively) with the Ward.D2 algorithm. t-SNE analysis revealed a high level of concordance (k = 4) with minor overlapping features between Group 3 and Group 4. Finally, we condensed the number of genes to 6 without significantly losing accuracy in classifying samples into SHH, WNT and non-SHH/non-WNT subgroups. Additionally, we found a relatively high frequency of WNT subgroup in our cohort, which requires further epidemiological studies. TLDA is a rapid, simple and cost-effective assay for classifying MB in low/middle income countries. A simplified method using six genes and restricting the final stratification into SHH, WNT and non-SHH/non-WNT appears to be a very interesting approach for rapid clinical decision-making.

Keywords: Medulloblastoma, Molecular subgroups, Brazilian cohort, Real-time PCR

* Correspondence: gavcruzeiro@gmail.com
†Luiz Gonzaga Tone and Elvis Terci Valera contributed equally to this work.
[1]Department of Pediatrics Ribeirão Preto Medical School, Hospital das Clínicas, University of São Paulo, Av. Bandeirantes 3900, Ribeirão Preto, São Paulo, Brazil
Full list of author information is available at the end of the article

Introduction

Medulloblastoma (MB) is the most common malignant brain tumor of children and adolescents, representing 20% of all pediatric brain tumors and is considered to be a complex disease from a genetic perspective [31]. Current consensus divides MB into four main molecular subgroups: SHH, WNT, Group 3 and Group 4. These subgroups have distinct transcriptional profiles, copy-number aberrations, somatic mutations and clinical outcomes [2, 18, 22–24].

The molecular subgroups of MB have been incorporated into risk stratification along with conventional biomarkers and preclinical models to evaluate novel targeted inhibitors and to substantiate further clinical trials [21, 23, 31]. The updated World Health Organization (WHO) classification of the central nervous system (CNS) acknowledges some of these molecular features as risk-stratification factors for MB [17]. Still, low and even middle-income countries cannot afford to routinely use these next-generation sequencing (NSG) platforms for MB molecular subgrouping. High costs related to these new technologies (e.g Illumina Methylation array 450 k and NanoString nCounter®) preclude their routine clinical application in most low-income Nations.

As an initial attempt to equate this subject, Kunder and colleagues [15] described a miRNA-based real-time PCR assay platform that performed subgroup assignment using a reduced set of 21 probes. However, analyses of Group 3 and Group 4 MB subgroups were not precisely discriminative when this approach was used and no algorithm accuracy was validated for their method. Similarly, Kaur and colleagues [12] published a simplified approach based on immunohistochemistry (IHC) and real time PCR (qPCR) methods for MB subgroup allocation [12]. However, overlapping IHC staining was observed between subgroups. More recently, complete datasets from cohort studies have become publicly available, allowing the validation for new molecular classification and comparing novel stratification proposals for gold standard NGS data. Accordingly, the validation of new algorithms seems to be critical considering their increasing genomic and molecular importance for therapeutic decisions [4, 7, 10, 16].

Here, we describe a low-cost and straightforward method for molecular allocation of MB patients. We hypothesized that a combination of qPCR with precise algorithms would be a useful, simple and potent tool for molecular assignment of MB tumors. We have optimized the number of genes to molecularly classify patients into four and three groups of interest for clinical management. We also present an elucidative algorithm for MB subgroup assignment, validating our approach and comparing our findings to data from 763 MB samples molecularly assigned through a robust integrative methodology (transcriptional, methylation and cytogenetic profiles) (GSE85217), as well as confirming our subgroup findings by Methylation array in a sample subset.

Methods

Study group

Ninety-two patients diagnosed with MB and treated at three Brazilian institutes were evaluated: 28 patients from the University Hospital, Ribeirão Preto Medical School, University of São Paulo (HC/FMRP-USP), 38 from The Boldrini Center of Children in Campinas São Paulo State, and 26 from the Medical School of São Paulo, University of São Paulo. In summary, 92 fresh-frozen MB tissue samples were microdissected by a single pathologist (F.S.P) in the Pathology Department (FMRP-USP). During microdissection, necrotic and normal tissues were removed from viable tumor areas. Patient age ranged from 1 to 24 years (median age = 7 years). Age groups at tumor diagnosis (clinical data from 88 MB patients out of 92 cases were available) were: infants (1–35 months) 11/88, children (36 months - 8 years) 38/88, and adolescents (9 to 17 years) 35/88. Tumors of young adults (age equals or above 18 + years) represented 4/88 of the case series. There was a slight male preponderance, with a male to female ratio of 1.30:1.0. From 92 samples, follow-up data of 80 patients were available and ranged from 1 to 168 months, with a median observation period of 41 months. Thirty-nine patients died because of the disease (DOD), 37 patients showed no evidence of disease (NED) at their last follow-up, and 4 patients (4.81%) died due to other unrelated events. Follow-up from a period of 1 to 5 years was available for 73% (59/80) of patients and 30% (24/80) were followed for more than 5 years. Two patients who lacked clinical information (named as “na” from HC-FMP/USP) were included in the heatmap for molecular assignment and were not considered for further analysis (clinical, demographic and survival (Additional file 1)). Patients’ clinical information, outcome, demographic and information on Methylation profile are presented in Fig. 1.

TaqMan low density array (TLDA)

Microdissected fresh frozen tumor tissues were submitted to RNA extraction using the RNAeasy kit (Qiagen). cDNA was synthetized in duplicate in a 25 μl reaction volume using 500 ng RNA from the High Capacity Kit (Thermo). After RT-PCR, 25 μl of DEPC water and 50 μl of Universal Master Mix (Life Technologies) were added at a ratio of 1:1. The TLDA plate layout was 31 + 1. The plate layout manufacturing control used was *GAPDH* and *18S* and the reference genes used were *TBP*, *HPRT* and *GUS-B*. Genes used for molecular assignment was retrieved from Codeset described by Northcott and colleagues

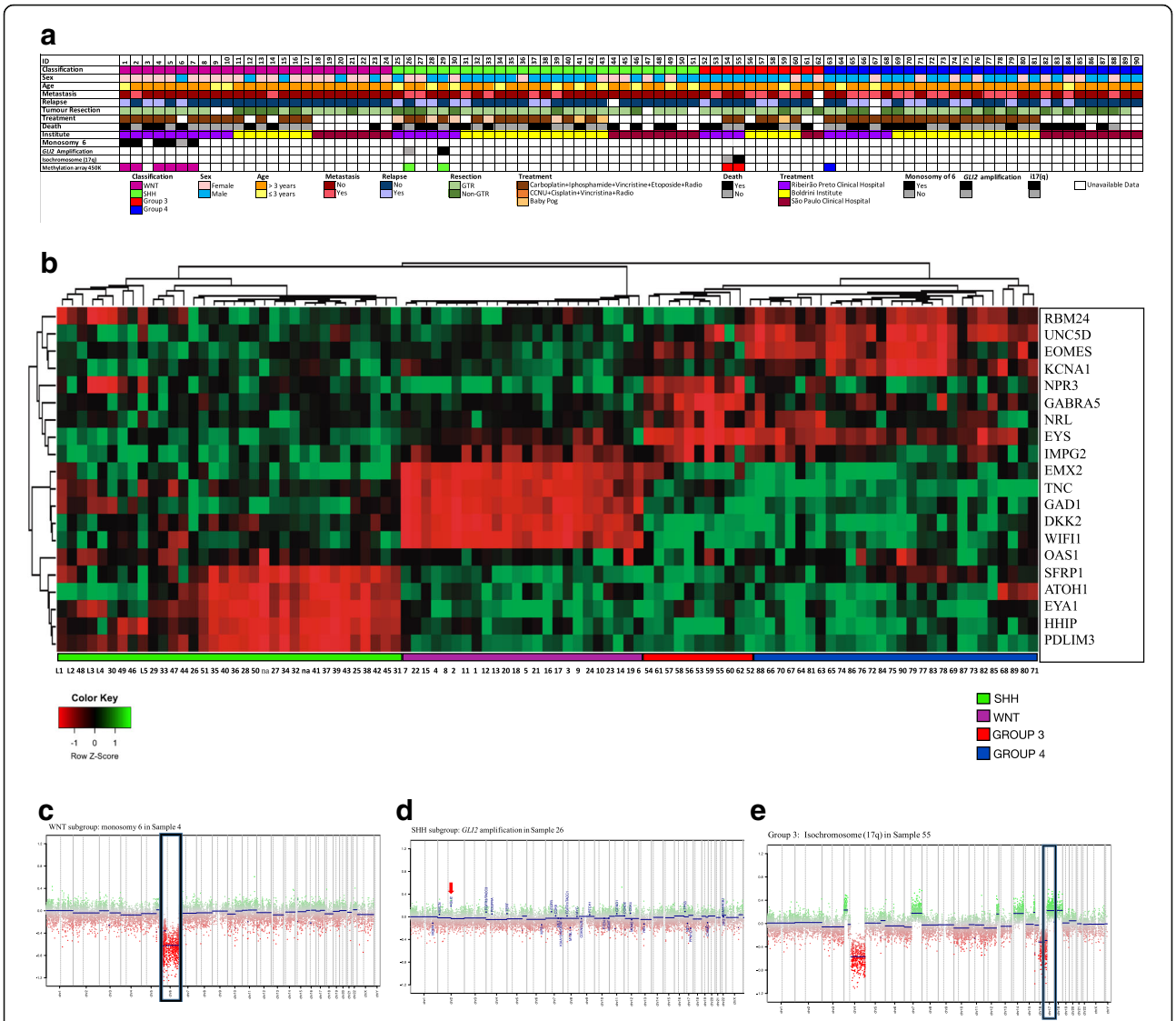

Fig. 1 a Clinical characteristics of MB patients ($n = 90$). *Classification* Molecular classification: WNT subgroup, SHH subgroup, Group 3 and Group 4 of patients. *Gender* (female and male). *Age at diagnosis* (below or above 3 years). *Metastasis* presence of metastasis at diagnosis (yes, no); *Relapse* presence of postoperative disease relapse (yes, no). *Tumor resection* (gross-total resection GTR; non-gross total resection non-GTR). *Treatment* treatment protocol (craniospinal radiotherapy plus carboplatin, ifosfamide, vincristine, etoposide; craniospinal radiotherapy plus CCNU (1-(2-chloroethyl)-3-cyclohexyl-1-nitrosourea), cisplatin, vincristine; Baby POG – Pediatric Oncology Group). *Death* if patient died (yes, no). *Institution* institute where patients received treatment, *Monosomy of chromosome 6* if patient bears this feature (yes, no), *GLI2 Amplification* if patients bears feature (yes, no), Isochromosome (17q) if patient bears feature (yes, no), *Methylation Array 450 K* Molecular assignment by methylation array of WNT (6), SHH (2), Group 3 (2) and Group 4 (1) samples. **b** Hierarchical unsupervised clustering of 92 primary MB into four molecular subgroups: SHH (green), WNT (purple), Group 3 (red) and Group 4 (blue). Pearson distance as Metric and average linkage as algorithm clustering. L1, L2, L3, L4 and L5 are represented as UW473, DAOY, UW402, UW228 and ONS-76 MB cell lines and "na" as samples tumors with unavailable data. **c** Copy number profile of sample 4 WNT subgroup (monosomy 6) (d) Copy number profile of sample 26 SHH Subgroup (Amplification of *GLI2*) (e) Copy number profile of sample 55 Group 3 (Isochromosome 17q)

[18] and probe set and manufacturer's code TaqMan probe are listed in (Additional file 2: Table S1). Four samples per plate were analyzed by conventional Real-Time PCR, with 1 μl Reaction Predesigned added to each well on Quant Studio 12 K flex (Thermo). The relative quantification (RQ) of each protein-coding gene compared with *TBP*, *HPRT* and *GUS-B* was determined by the comparative cycle threshold (Ct) method, where RQ 1/4 22(Ct Gene 2Ct Ref) × 100.

Molecular assignment of MB samples

Codeset genes expression analysis was used to generate a pairwise distance matrix. Additionally, MB cell lines

previously assigned to the SHH subgroup: DAOY, UW228, ONS-76 [9, 18, 29, 30] and the UW402 and UW473 (no subgroup information) were included in the analysis.

For molecular subgroup assignment, unsupervised hierarchical clustering was performed by Pearson distance correlation followed by an average-linkage algorithm. Delta Ct values were used during analysis and a Heatmap was generated using the Expression Suite® software (Life Technologies). A total of 763 MB samples from the study of Cavalli and colleagues [1] (GSE85217) in the R environment were analyzed for algorithm validation and heatmap comparison.

Molecular assignment of MB samples by methylation array and copy number profiling

In order to assess concordance between TLDA assay and the gold standard Illumina 450 K Methylation array, DNA was extracted from 11 fresh frozen MB tumors and 250 ng were processed for genome-wide DNA methylation analysis using the Illumina HumanMethylation450 BeadChip (450 k). t-SNE analysis (t-Distributed Stochastic Neighbor Embedding, Rtsne package version 0.11) was performed and MB samples were randomly tested along with 390 MB reference samples molecularly assigned in Capper and colleagues study (GSE109381) [1]. MB samples were further submitted to DNA methylation class prediction and calibrated random forest class prediction scores were generated using classifier version 11. b4 based on the analysis of 10,000 CpG sites present in the 450 k. For molecular subgrouping based on methylation class, an optimal calibrated score threshold was defined as ≥0.5 for a sufficient prediction as long as all family member scores add up to a total score of ≥0.9. Additionally, copy-number variations (CNV) analysis was performed using the 'conumee' R package in bioconductor. Briefly, all CpGs are represented by a methylation probe and unmethylated probe. For CNVs identification, the methylated and unmethylated signal intensities are added together and a ratio is formed against a healthy reference (e.g normal cerebellum tissue) that bears a flat genome. Finally, the relative copy-number is plotted in the corresponding area of chromosomal location. The automatic score is verified by manual curation of the respective loci for each individual profile [8, 29].

Bioinformatic analysis

The R language and environment for statistical computing and graphics was used for bioinformatic analysis. The ComplexHeatmap and circlize packages were used for Heatmap generation [5, 6] and the ggplot2package [26, 32] was used for graphics generation. Rtsne [14, 10] was used for the visualization of t-Distributed Stochastic Neighbor Embedding (t-SNE) and the NbClust and Factoextra packages [3, 11] were used to point out the best number of clusters and to visualize the results. Pearson correlation was used as a distance parameter and we selected the clustering algorithms Ward.D2 or Average linkage. To perform t-SNE, a method for constructing a low dimensional embedding of high-dimensional data, distances or similarities, we used the default parameters, setting only the perplexity parameter at 30, with 5–50 being the typical and recommended range for robust analysis. We then used NbClust with default parameters to find the best number of cluster and to visualize the results.

Statistical analysis

The SPSS version 20 software (SPSS Inc., Chicago, IL, USA) was used for statistical analysis. Descriptive statistics and frequency distributions were calculated for all variables; the chi-square test and Fisher's exact test were applied to explore association between variables. Patients' event-free survival (EFS) rates were evaluated by Kaplan-Meier curves and the log-rank test, considering relapse/death as the event. *P*-values < 0.05 were considered to be significant. [* < 0.05, ** < 0.01, *** < 0.001].

Results

TLDA accurately assigned most of the MB samples as WNT, SHH, Group 3 and Group 4

As previously described [13, 15, 19] the SHH MB subgroup was defined according to the expression of *PDLIM3*, *EYA1*, *HHIP*, *ATOH1*, *SFRP1*, with *EYA1 HHIP* and *SFRP1* being the genes most expressed in 27 SHH MB patients (30%). The WNT subgroup in 24 patients (27%) was represented by the *OAS1*, *WIFI1*, *DKK2*, *TNC*, *GAD1* and *EMX2* genes, with *WIFI1*, *DKK2* and *EMX2* being the most expressed genes in this subgroup. *IMPG2*, *EYS*, *NRL*, *GABRA5* expression was used to assign 11 MB samples (12%) to Group 3, and *GABRA5* and *NPR3* expression was the most specific subgroup compared to Group 4. Twenty-eight Group 4 MBs (31%) were assigned using *KCNA1*, *EOMES*, *UNC5D*, and *RBM24* expression, with *KCNA1* and *RBM24* being the most specific subgroup compared to Group 3. Specifically, *TNC* showed higher expression in WNT subgroup and average expression in SHH MB. The *EYS* gene was differentially expressed in Group 3 and Group 4 (Fig. 1b). DAOY, UW228 and ONS-76 cell lines were confirmed to belong to SHH subgroup. UW473 and UW402 were subgrouped as SHH MB as well.

Methylation and copy number profiling of MBs using illumina methylation array 450 K showed high concordance with TLDA

In order to validate our method, DNA available of 11 randomized MB patients were submitted to Methylation array 450 K (Copy number profile available in Fig. 1). We found a high concordance between Methylation

array 450 K and TLDA for molecular assignment of MBs. The t-SNE analysis of eleven MB samples along with 390 MB samples (GSE109381) showed high concordance with TLDA method, being all samples assigned in the same molecular subgroup (Additional file 3: Figure S1). The DNA methylation class prediction and calibrated random forest class prediction scores identified 6 WNT MBs, 2 SHH MBs, two Group 3 MBs and one Group 4 MB (Additional file 4: Table S2). Additionally, copy number profiling identified monosomy in chromosome 6 in WNT subgroup ($n = 5$), *GLI2* amplification in SHH ($n = 1$) and I (17q) for Group 3 MBs ($n = 1$) (Fig. 1c, d and e respectively).

T-SNE analysis revealed concordance between the Brazilian cohort and the validation cohort and highlighted overlapping features of group 3 and group 4

t-SNE analysis was performed to visualize clustering features of molecular subgroups in perplexity index of 30. We found four subgroups in the Brazilian cohort study, with Group 3 and Group 4 bearing overlapping features ($k = 4$). To validate this analysis, the t-SNE algorithm was also applied to the validation cohort of 763 MB samples and the data obtained showed the same behavior ($k = 4$) (Fig. 2a and b).

Average linkage and Ward.D2 are robust algorithms for subgroup assignment of MB

In order to compare the clusterization feature algorithms Ward and Average-linkage we applied our TLDA approach to a validation cohort of 763 pre-classified MB samples submitted to an integrative methodology composed of transcriptional, methylation profile and cytogenetic features. Interestingly, we found both Average-linkage and Ward.D2 to be feasible algorithms for MB subgroup assignment using transcriptional data alone. The Average-linkage algorithm successfully assigned 221 of 223 SHH MB samples (99.10% accuracy), 66 from 70 WNT MB samples (94.29% of accuracy), 133 from 144 MB Group 3 MB samples (92.36% accuracy), and 311 from 326 Group 4 MB samples (95.40% accuracy). Equally, the Ward.D2 algorithm successfully assigned 216 of 223 SHH MB samples (97.31% accuracy), 68 from 70 WNT MB samples (97.14% accuracy), 128 from 144 MB Group 3 MB samples (88.89% accuracy), and 317 from 326 Group 4 MB samples (97.24% accuracy). (Fig. 3a and b) (Table 1).

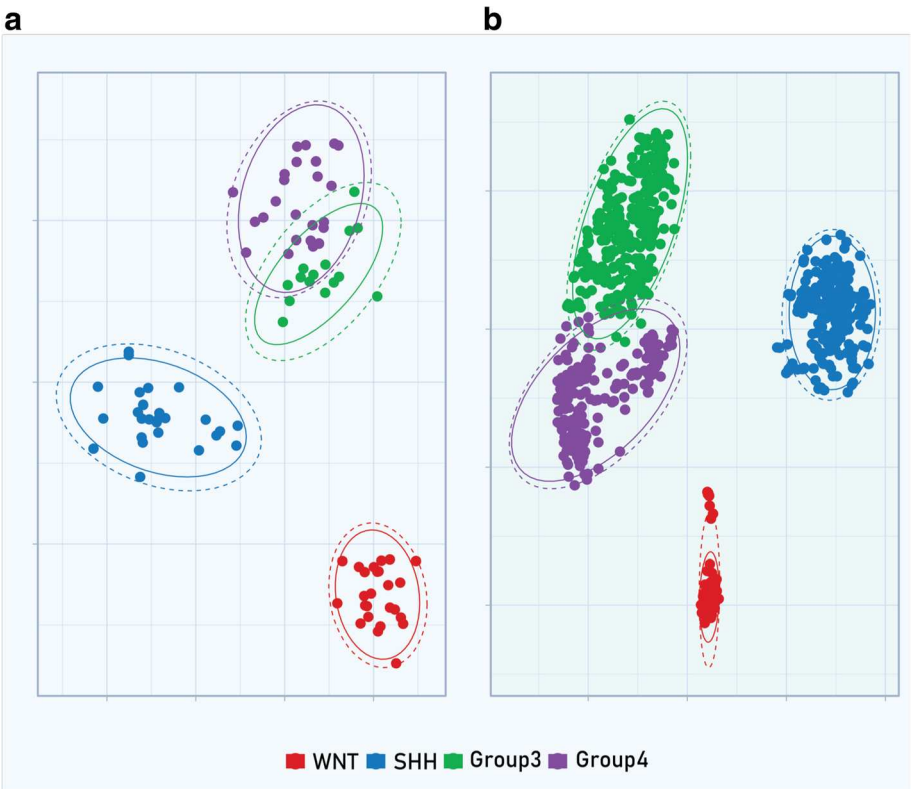

Fig. 2 **a** Two-dimensional representation of pairwise sample correlations of twenty TaqMan expression assay probes (Additional file: Table S1) in 92 MB Brazilian samples by t-Distributed Stochastic Neighbor Embedding. **b** Two-dimensional representation of pairwise sample correlation of the same gene set represented in (**a**) using Microarray probes in 763 MB samples from GSE85217 by t-Distributed Stochastic Neighbor Embedding

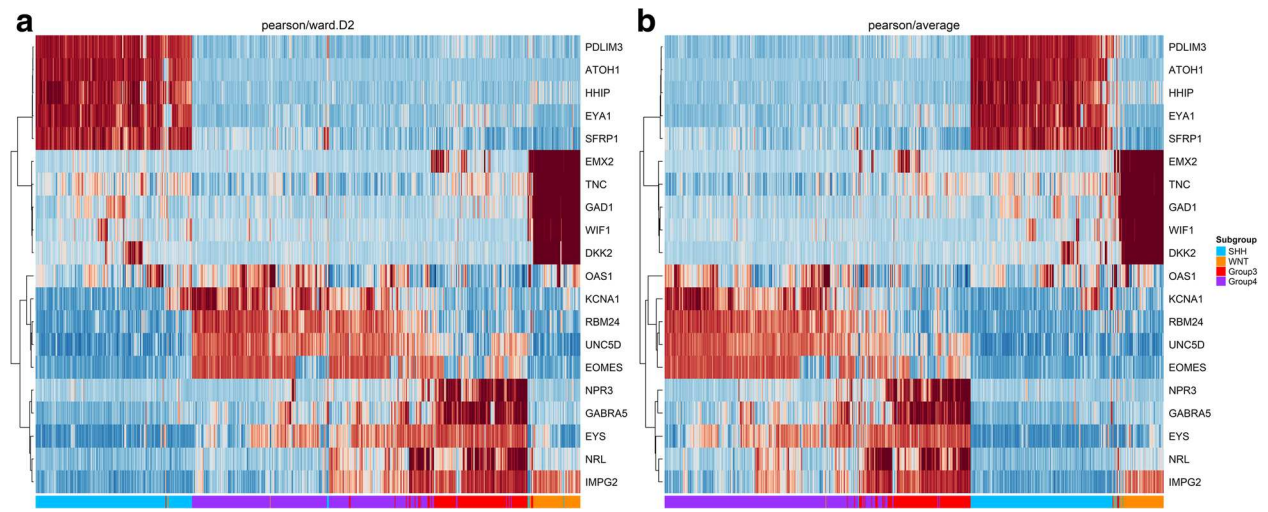

Fig. 3 Hierarchical unsupervised clustering of previously classified 763 primary MB in GSE85217 study: SHH (blue), WNT (orange), Group 3 (red) and Group 4 (purple) Pearson distance as Metric was utilized in both heatmaps. **a** Clustering using the Ward.D2 algorithm. **b** Clustering using the Average linkage algorithm

Analysis of *SFRP1*, *HHIP*, *EYA1*, *WIFI1*, *EMX2* and *DKK2* expression potentially discriminated between SHH, WNT from non-SHH/non-WNT MB

We further performed expression analysis of six key genes that bear a positive signature for SHH (*SFRP1*, *HHIP*, *EYA1*) and WNT (*WIFI1*, *EMX2* and *DKK2*) using Pearson as distance measurement and Ward.D2 or Average linkage as clustering algorithms. We found that the first cluster was characterized by a differential expression of *SFRP1*, *HHIP* and *EYA1*, which represent the SHH subgroup. Another cluster that differentially carried expression of *WIFI1*, *EMX2* and *DKK2* represented the WNT subgroup. The third cluster, which carried very low levels or lacked expression of the six genes, was assigned as N-WNT/N-SHH. Similarly, in the validation cohort of 763 samples, we identified the same behavior, indicating the presence of three main clusters (Fig. 4a and b). Using t-SNE analysis, we observed the same consistent assignment of MB samples to 3 main clusters ($k = 3$), with a minor overlap of clusters N-SHH/N-WNT and SHH (Fig. 5a and b) (Additional file 5: Figure S2a and b). The accuracy of subgroup assignment using the set of six genes is showed in Table 2a and b.

Discussion

In the present study, differential expression analysis of 20 genes from the CodeSet described by Northcott and colleagues [19] by TDLA approach permitted us to molecularly assign a cohort of 92 MB patients to the four major MB subgroups. Additionally, we validated the same gene set in a cohort of 763 MB patients from the GSE85217 reference study, which applied the integrative-clustering method to molecularly classify MB samples. The WNT and SHH subgroups were robustly identified since they formed a solid and concise cluster generated by the Average-linkage or Ward.D2 algorithms and confirmed by t-SNE analysis. In agreement, similar patterns were detected using GSE85217 data analysis. We demonstrated that assessment of the transcription profile is not sufficient to completely discriminate all Group 3 MB from Group 4 MB since a minority of these patients share transcription and common molecular features [10, 12, 15, 18].

Next, in order to exam the concordance of our TDLA approach with NGS subgrouping for MB we validated molecular assignment of 11 MBs samples by Methylation Array 450 K. We found a high frequency of monosomy in chromosome 6 within WNT (5 out of 6) subgroup corroborating with previous studies [2, 8, 13,

Table 1 Comparison of algorithm accuracy in the GSE85217 study ($n = 763$). Misassignment is defined as patients who were incorrectly subgrouped

	Misassignment		Efficiency	
Subgroup	**Average linkage**	**Ward.D2**	**Average-linkage**	**Ward.D2**
SHH (n=223)	2	6	99.10%	97.31%
WNT (n=70)	4	2	94.29%	97.14%
Group 3 (n=144)	11	16	92.36%	88.89%
Group 4 (n=326)	15	9	95.40%	97.23%

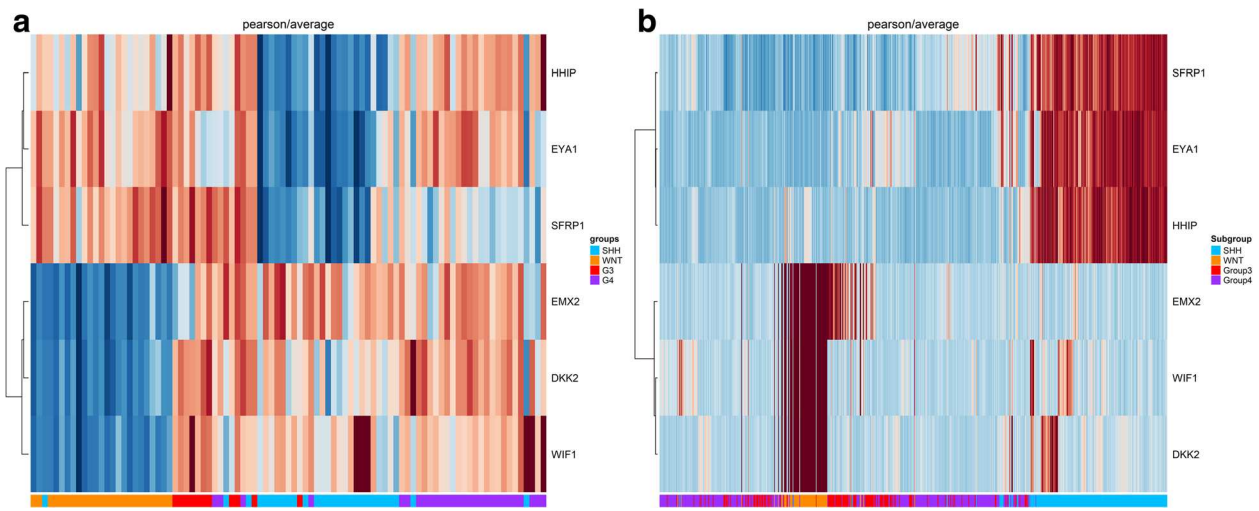

Fig. 4 Hierarchical unsupervised clustering using *HHIP*, *EYA1*, *SFRP1*, *EMX2*, *DKK2*, *WIFI1* **a** 92 MB samples from Brazilian cohort and **b** 763 MB samples from GSE85217. SHH (blue), WNT (orange), Group 3 (red) and Group 4 (purple)

28]. In one SHH MB samples evaluated by Methylation array we identified *GLI2* amplification. For Group 3, one MB specimen bears isochromosome 17q, a reliable marker for this subgroup [28] (Fig. 2B). Only one sample for group 4 was identified, and it also clustered to group 4 by TLDA method accordingly. Full concordance between eleven MB samples by NGS and TDLA was observed. Despite only a small set of samples was assessed, the results from NGS data support our molecular assignment provided by TLDA [2, 8, 13, 28].

In the present study, we found 27% of WNT MBs (Additional file 6: Figure S3a-S3d and Additional file 7: Figure S4). Although this is a high frequency when compared to studies performed in North American and

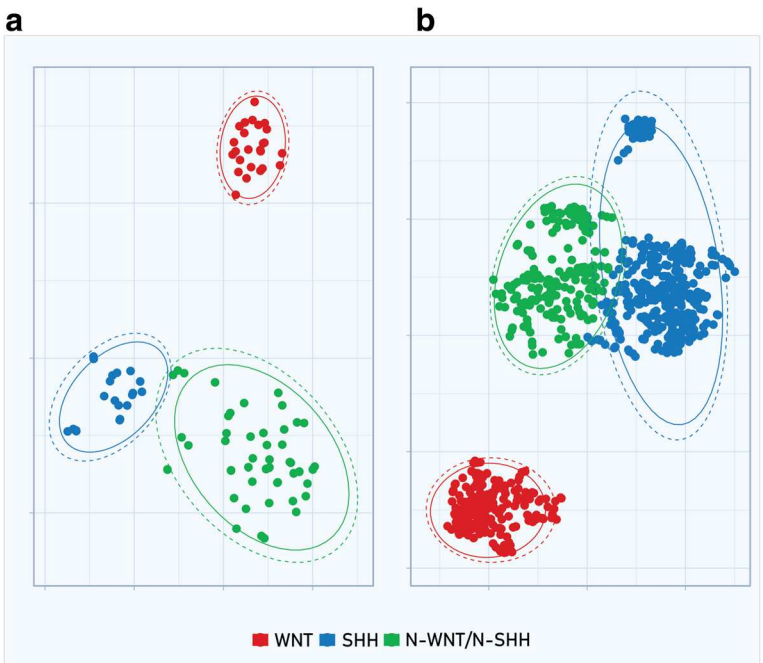

Fig. 5 **a** Two-dimensional representation of pairwise sample correlations of 6 TaqMan expression assay probes (*SFRP1*, *HHIP*, *EYA1*, *WIFI1*, *EMX2* and *DKK2)* in 92 MB Brazilian samples by t-Distributed Stochastic Neighbor Embedding. **b** Two- dimensional representation of pairwise sample correlation of the same gene set represented in (**a**), although using Microarray probes of 763 MB samples from GSE85217 by t-Distributed Stochastic Neighbor Embedding

Table 2 Comparison of algorithm accuracy using 6 genes to assign WNT and SHH alone. (**a**) Study GSE85217 ($n = 293$). (**b**) Brazilian cohort ($n = 51$). Misassignment is defined as patients who were incorrectly subgrouped

a

	Misassignment		*Accuracy*	
Subgroups	**Average-linkage**	**Ward.D2**	**Average-linkage**	**Ward.D2**
SHH (n=223)	17	19	92.37%	91.47%
WNT (n=70)	12	11	82.85%	84.28%

b

	Misassignment		*Accuracy*	
Subgroups	**Average-linkage**	**Ward.D2**	**Average-linkage**	**Ward.D2**
SHH (n=27)	4	4	85.18%	85.18%
WNT (n=24)	0	0	100%	100%

European continents [19], Kunder and colleagues [15] reported 24% of WNT MBs in an Indian cohort. Moreover, pediatric neoplasms subtypes vary in frequency depending on the genetic population background (i.e: high frequency of Promyelocitic Leukemia in Latin America) [20, 25]. Interestingly, we found 2 cases of desmoplastic and 1 LCA in WNT MBs. Besides it is unlikely to find desmosplastic histological variants in WNT MBs, our data are supported by other studies [27]. In summary, these epidemiological facts highlight the urge for a reliable, feasible and low-cost method to perform molecular assignment of MBs in low and middle-income countries.

The average-linkage and Ward.D2 algorithms were assessed regarding their clustering features and subgroup assignment. In the GSE85217 study conducted on 763 MB patients, average-linkage provided better accuracy for SHH and Group 3 assignment compared to the Ward.D2 method. However, Ward.D2 was able to accurately classify WNT and Group 4 tumors. Interestingly, the pick of an accurate clustering algorithm may be subgroup specific. However, it is very important to understand the limitations of transcriptional data and information that can be extracted from a single feature such as gene expression [2, 27, 33].

Indeed, as reported by Cavalli and colleagues [2], the gold standard method for subgroup assignment is the assessment of the molecular features of the patient (transcription profile, methylation profile, cytogenetic profile) along with clinical information. However, in low-income countries, most molecular techniques are onerous for application to daily clinical practice. Using expression analysis of a gene set, algorithm assessment and bioinformatic analysis, we sought to identify the minimal number of genes needed to molecularly classify MB as WNT, SHH and non-SHH /non-WNT. In our study, by using a set of six differentially expressed genes we were able to distinguish SHH and WNT from non-WNT/non-SHH without s significant loss of accuracy. Both the Average-linkage and Ward.D2 algorithms conserved 100% accuracy for assignment to the WNT subgroup, with a decline to 85.18% for the SHH subgroup. As shown in the t-SNE map, there was a minor overlapping of samples of the non-SHH/non-WNT cluster with those of the SHH cluster. Additionally, we found high concordance between our data set and GSE85217, with 100% accuracy for the WNT subgroup and 86% accuracy for SHH. These results shed new light on a potential method for low-income countries based on a simple and feasible technique such as qPCR along with six probe/primer pairs plus reference genes with implementation of an approach recently described by Gómes and colleagues [3]. Their method fully discriminates between Group 3 and Group 4 based on the methylation status of 5 CpG's, which is feasible for the real-time PCR platform through High Resolution Melting technology, and shall improve the molecular assignment [26].

Northcott et al. described a molecular classification method for MB that relies on the NanoString nCounter System. Besides the high accuracy of the method (~ 98%), the average cost is estimated at 60.00 USD per sample and the method takes 3–4 days to perform bioinformatic analysis [19]. The same method was reproduced by Leal and colleagues [16]; however, due to the high cost of the equipment (287,817.60 USD – average price in South America; 2018), it is challenging for most low-income countries to apply this method to clinical routine. Kaur et al. proposed a minimal panel comprising a combination of IHC antibodies and FISH probes to classify MB, with an estimate cost around 150.00 to 250.00 USD (average) per sample [12]. Although feasible, their approach does not seem to be as cost-effective as other methods and IHC analysis remains challenging due to different antibody batches and inter-observer consistency [19]. More recently, the minimal methylation classifier (MIMIC) was described as a highly efficient methodology that might be superior to Illumina 450 K and Methylation EPIC array for MB molecular assignment regarding feasibility for clinical routines; however, the average cost per sample with this approach is around 200.00 USD [26] and requires the acquisition of a MALDI-TOF mass-spectrometer (approximately 150,000.00 USD), along with a conventional PCR device. Our method using TLDA has an estimated cost of 70.00 USD per sample (including reagents, primers and laboratory implements). The equipment necessary to run TLDA costs about 92,600.00 USD and complete data analysis is ready within one working day. Moreover, when we condensed the number of studied genes to six (a set of TaqMan probes for: *SFRP1*, *HHIP*, *EYA1*, *WIFI1*, *EMX2* and *DKK2* along with reference genes *HPRT* and *Gus-ß*), the cost of molecular assignment to the WNT, SHH and N-SHH/N-WNT MB subgroups dropped to 26.82 USD per sample. Also, the real time-PCR (30,000.00 USD) platform is relatively inexpensive and commonly available in most hospitals due to its ample use for other routine laboratory applications. Finally, another

advantage of the qPCR method is that it does not require batched minimal number of samples per run, being readily available to run single tumor samples upon arrival at the laboratory.

Conclusions

In conclusion, we have developed a simplified approach and validated TDLA method in random samples by Methylation Array 450 K. In addition, our findings were challenged at a large cohort study GSE85217 through accurate algorithms for molecular assignment of MB. The proposed assay is cost-effective and discriminates most of SHH, WNT, non-SHH/non-WNT tumors. The TLDA method for MB subgroup stratification might be an affordable tool to be used to drive therapies in low-income countries. Moreover, it may also be an important approach for prompt classification and decision-making algorithms in MB, before NGS data analysis become available.

Additional files

Additional file 1: Demographic analysis from our cohort study. Corresponding figures are Additional file 6: Figure S3a, S3b, S3c and S3d. (DOCX 13 kb)

Additional file 2: Table S1. TaqMan probes from the gene set used for MB molecular assignment. (CSV 787 bytes)

Additional file 3: Figure S1. t-SNE map show molecular assignment by Methylation array 450 K of 11 MB samples from our study along with 390 MB samples from GSE109381. (PDF 61 kb)

Additional file 4: Table S2. Report of DNA class prediction classifier using Random forest class prediction scores (classifier version 11. b4). (CSV 564 bytes)

Additional file 5: Figure S2. Comparison of clustering algorithms in our study ($n = 92$) with 6 genes *HHIP*, *EYA1*, *SFRP1*, *EMX2*, *DKK2*, *WIFI1*. (**a**) Ward.D2 algorithms (**b**) Average-linkage algorithms. (PDF 1980 kb)

Additional file 6: Figure S3. (**a**) Demographic distribution of the 4 molecular subgroups in the present cohort; (**b**) subgroup distribution with respect to age at diagnosis; (**c**) gender; (**d**) histological variants. The numbers indicate the sum of tumors in each category. (PDF 85 kb)

Additional file 7: Figure S4. Overall survival of molecular subgroups ($n = 80$). (PDF 62 kb)

Acknowledgements

This study was funded by FAPESP (grant numbers 2013/12006-3; 2014/19976-0; 2014/20341-0; 04/12133-6 and 2013/02162-8). We want to thank Dr. David T. W. Jones from DKFZ who kindly helped with methylation profile analysis and medulloblastoma molecular assignment. Finally, we want to thank and acknowledge patients and families affected by medulloblastoma for their generous contributions to these studies. We also had support of FAEPA (Fundação de Amparo ao Ensino, Pesquisa e Assistência do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto-USP) for publication.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding authors upon request.

Authors' contributions

GAVC planned and conducted all experiments, drafted and critically read the manuscript, KBS helped to design the study and critically read the manuscript, CAOB performed in silico analysis in the R environment, MB helped with Methylation Array 450 K and critically read the manuscript, DS helped with Methylation Array 450 K and copy number profiling, RCPL conducted the overall survival analysis, TAM helped to design the study and graphics, MBM assembled and helped to organize the datasheet, VSS helped to design the study and critically read the manuscript, FPS performed the histopathologic analysis, RGPQ and SMOS helped with sample processing, SMKN provided patient samples and critically read the manuscript, CAS critically read the manuscript, LGT designed the study and critically read the manuscript. ETV helped with clinical data and critically read the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This research was submitted to and approved by the HC/FMRP-USP Research Ethics Committee (CAAE n° 37,206,114.1.0000.5440) n°15,509/2016. All samples were obtained after receiving informed consent from all participants included in the study.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Author details

[1]Department of Pediatrics Ribeirão Preto Medical School, Hospital das Clínicas, University of São Paulo, Av. Bandeirantes 3900, Ribeirão Preto, São Paulo, Brazil. [2]Department of Genetics, Ribeirão Preto Medical School, University of São Paulo, Av. Bandeirantes 3900, Ribeirão Preto, São Paulo, Brazil. [3]Department of Oncology, Children's Research Center, Neuro-Oncology group, University Children's Hospital Zürich, August-Forel Strasse 1, CH-8008 Zürich, Switzerland. [4]Pediatric Glioma Research Group, Hopp Children's Cancer Center at the NCT Heidelberg (KiTZ) and German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. [5]Department of Pathology, University of São Paulo, Av. Bandeirantes 3900, Ribeirão Preto, SP 14049-900, Brazil. [6]Division of Pediatric Neurosurgery, Department of Surgery and Anatomy, Ribeirão Preto Medical School, Hospital das Clínicas, University of São Paulo, Av.Bandeirantes 3900, Ribeirão Preto, São Paulo, Brazil. [7]Boldrini Centre of Children, University of Campinas-UNICAMP, Campinas, SP, Brazil. [8]Department of Neurology, Faculty of Medicine, University of São Paulo, São Paulo, Brazil.

Received: 22 January 2019 Accepted: 18 February 2019
Published online: 04 March 2019

References

1. Capper D, Jones DTW, Sill M, Hovestadt V, Schrimpf D, Sturm D et al (2018) DNA methylation-based classification of central nervous system tumours. Nature 555(7697):469–474. https://doi.org/10.1038/nature26000 Epub 2018 Mar 14. PubMed PMID: 29539639; PubMed Central PMCID: PMC6093218
2. Cavalli FMG, Remke M, Rampasek L, Peacock J, Shih DJH, Luu B et al (2017) Intertumoral heterogeneity within medulloblastoma subgroups. Cancer Cell 31(6): 737–754. https://doi.org/10.1016/j.ccell.2017.05.005 e6. PubMed PMID: 28609654
3. Charrad M (2012) NbClust Package: finding the relevant number of clusters in a dataset. UseR!
4. Gómez S, Garrido-Garcia A, Garcia-Gerique L, Lemos I, Suñol M, de Torres C, Kulis M et al (2018) A novel method for rapid molecular subgrouping of Medulloblastoma. Clin Cancer Res 24(6):1355–1363. https://doi.org/10.1158/1078-0432.CCR-17-2243 Epub 2018 Jan 19. PubMed PMID: 29351917
5. Gu Z, Eils R, Schlesner M (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 32(18):2847–2849. https://doi.org/10.1093/bioinformatics/btw313 Epub 2016 May 20. PubMed PMID: 27207943
6. Gu Z, Gu L, Eils R, Schlesner M, Brors B (2014) Circlize implements and enhances circular visualization in R. Bioinformatics 30(19):2811–2812. https://doi.org/10.1093/bioinformatics/btu393 Epub 2014 Jun 14. PubMed PMID: 24930139

7. Holgado BL, Guerreiro Stucklin A, Garzia L, Daniels C, Taylor MD (2017) Tailoring Medulloblastoma treatment through genomics: making a change, one subgroup at a time. Annu Rev Genomics Hum Genet 18:143–166. https://doi.org/10.1146/annurev-genom-091416-035434 Epub (2017) May 5. Review. PubMed PMID: 28475368
8. Hovestadt V, Remke M, Kool M, Pietsch T, Northcott PA, Fischer R et al (2013) Robust molecular subgrouping and copy-number profiling of medulloblastoma from small amounts of archival tumour material using high-density DNA methylation arrays. Acta Neuropathol 125(6):913–916. https://doi.org/10.1007/s00401-013-1126-5. Epub 2013 May14 PubMed PMID: 23670100; PubMed Central PMCID: PMC3661908
9. Ivanov DP, Coyle B, Walker DA, Grabowska AM (2016) In vitro models of medulloblastoma: choosing the right tool for the job. J Biotechnol 236:10–25. https://doi.org/10.1016/j.jbiotec.2016.07.028 Epub 2016 Aug 3. Review. PubMed PMID: 27498314
10. Karkucińska-Więckowska A, Kaleta M, Drogosiewicz M, Perek-Polnik M, Krętowski A, Cukrowska B et al (2018, 2018) Medulloblastoma with transitional features between Group 3 and Group 4 is associated with good prognosis. J Neurooncol. https://doi.org/10.1007/s11060-018-2797-5 [Epub ahead of print] PubMed PMID: 29427151
11. Kassambra A, Mundt F (2016) Factoextra: extract and visualize the results of multivariate data analyses. R package version, v. 1, n. 3, 2016
12. Kaur K, Kakkar A, Kumar A, Mallick S, Julka PK, Gupta D et al (2016) Integrating molecular subclassification of medulloblastomas into routine clinical practice: a simplified approach. Brain Pathol 26(3):334–343. https://doi.org/10.1111/bpa.12293 Epub 2015 Sep 9. PubMed PMID: 26222673
13. Korshunov A, Chavez L, Northcott PA, Sharma T, Ryzhova M, Jones DTW et al (2017) DNA-methylation profiling discloses significant advantages over NanoString method for molecular classification of medulloblastoma. Acta Neuropathol 134(6):965–967. https://doi.org/10.1007/s00401-017-1776-9 Epub 2017 Oct 13. PubMed PMID: 29027579
14. Krijthek, J. H (2015). Rtsne: T-distributed stochastic neighbor embedding using Barnes-Hut implementation. R package version 0.13. https://github.com/jkrijthe/Rtsne.
15. Kunder R, Jalali R, Sridhar E, Moiyadi A, Goel N, Goel A et al (2013) Real-time PCR assay based on the differential expression of microRNAs and protein-coding genes for molecular classification of formalin-fixed paraffin embedded medulloblastomas. Neuro-Oncology 15(12):1644–1651. https://doi.org/10.1093/neuonc/not123
16. Leal LF, Evangelista AF, de Paula FE, Caravina Almeida G, Carloni AC, Saggioro F et al (2018) Reproducibility of the NanoString 22-gene molecular subgroup assay for improved prognostic prediction of medulloblastoma. Neuropathology. https://doi.org/10.1111/neup.12508 [Epub ahead of print] PubMed PMID: 30155928
17. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK et al (2016) The 2016 World Health Organization classification of tumors of the central nervous system: a summary. Acta Neuropathol 131(6):803–820. https://doi.org/10.1007/s00401-016-1545-1 Epub 2016 May 9. Review. PubMed PMID: 27157931
18. Northcott PA, Jones DT, Kool M, Robinson GW, Gilbertson RJ, Cho YJ et al (2012) Medulloblastomics: the end of the beginning. Nat Rev Cancer 12:818–834
19. Northcott PA, Shih DJ, Remke M, Cho YJ, Kool M, Hawkins C et al (2012) Rapid, reliable, and reproducible molecular sub-grouping of clinical medulloblastoma samples. Acta Neuropathol 123(4):615–626. https://doi.org/10.1007/s00401-011-0899-7 Epub 2011 Nov 6. PubMed PMID: 22057785; PubMed Central PMCID: PMC3306784
20. Otero JC, Santillana S, Fereyros G (1996) High frequency of acute promyelocytic leukemia among Latinos with acute myeloid leukemia. Blood 88(1):377 PubMed PMID: 8704198
21. Pei Y, Liu KW, Wang J, Garancher A, Tao R, Esparza LA et al (2016) HDAC and PI3K antagonists cooperate to inhibit growth of MYC-driven medulloblastoma. Cancer Cell 29:311–323
22. Ramaswamy V, Remke M, Adamski J, Bartels U, Tabori U, Wang X et al (2016) Medulloblastoma subgroup-specific outcomes in irradiated children: who are the true high-risk patients? Neuro-Oncology 18:291–297
23. Ramaswamy V, Remke M, Bouffet E, Bailey S, Clifford SC, Doz F et al (2016) Risk stratification of childhood medulloblastoma in the molecular era: the current consensus. Acta Neuropathol 131:821–831
24. Ramaswamy V, Remke M, Bouffet E, Faria CC, Perreault S, Cho YJ et al (2013) Recurrence pat- terns across medulloblastoma subgroups: an integrated clinical and molecular analysis. Lancet Oncol 14:1200–1207
25. Rego EM, Jácomo RH (2011) Epidemiology and treatment of acute promyelocytic leukemia in latin america. Mediterr J Hematol Infect Dis 3(1): e2011049. https://doi.org/10.4084/MJHID.2011.049 Epub 2011 Oct 24. PubMed PMID: 22110899; PubMed Central PMCID: PMC3219651
26. Schwalbe EC, Hicks D, Rafiee G, Bashton M, Gohlke H, Enshaei A et al (2017) Minimal methylation classifier (MIMIC): A novel method for derivation and rapid diagnostic detection of disease-associated DNA methylation signatures. Sci Rep 7(1):13421. https://doi.org/10.1038/s41598-017-13644-1 PubMed PMID: 29044166; PubMed Central PMCID: PMC5647382
27. Schwalbe EC, Lindsey JC, Nakjang S, Crosier S, Smith AJ, Hicks D et al (2017) Novel molecular subgroups for clinical classification and outcome prediction in childhood medulloblastoma: a cohort study. Lancet Oncol 18(7):958–971. https://doi.org/10.1016/S1470-2045(17)30243-7.Epub PubMed PMID: 28545823; PubMed Central PMCID: PMC5489698
28. Shih DJ, Northcott PA, Remke M, Korshunov A, Ramaswamy V, Kool M et al (2014) Cytogenetic prognostication within medulloblastoma subgroups. J Clin Oncol 32(9):886–896. https://doi.org/10.1200/JCO.2013.50.9539 Epub 2014 Feb 3. PubMed PMID: 24493713; PubMed Central PMCID: PMC3948094
29. Sturm D, Witt H, Hovestadt V, Khuong-Quang DA, Jones DT, Konermann C (2012) Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. Cancer Cell 22(4):425–437. https://doi.org/10.1016/j.ccr.2012.08.024 PubMed PMID: 23079654
30. Triscott J, Lee C, Foster C, Manoranjan B, Pambid MR, Berns R et al (2013) Personalizing the treatment of pediatric medulloblastoma: polo-like kinase 1 as a molecular target in high-risk children. Cancer Res 73(22):6734–6744. https://doi.org/10.1158/0008-5472.CAN-12-4331 Epub 2013 Sep 9. PubMed PMID: 24019381
31. Wang J, Garancher A, Ramaswamy V, Wechsler-Reya RJ (2018) Medulloblastoma: from molecular subgroups to molecular targeted therapies. Annu Rev Neurosci 41:207–232. https://doi.org/10.1146/annurev-neuro-070815-013838 Epub 2018 Apr 11. PubMed PMID: 29641939
32. Wickham H (2016) ggplot2: elegant graphics for data analysis. Springer; https://www.springer.com/us/book/9780387981413
33. Xu R, Wunsch D (2005) 2nd survey of clustering algorithms. IEEE Trans Neural Netw 16(3):645–678 Review. PubMed PMID: 15940994